

5

DESCRIPTIONNOVEL THERMOSTABLE PROTEOLYTIC ENZYMES  
AND USES THEREOF IN PEPTIDE AND PROTEIN SYNTHESIS

10

Cross-Reference to a Related Application

This application is a continuation of USSN 10/379,952, filed March 4, 2003, which is a divisional of USSN 09/404,031, filed September 23, 1999, now U.S. Patent No. 6,573,065, which is a divisional of USSN 09/197,945, filed November 23, 1998, now U.S. Patent No. 6,143,517, which is a continuation of USSN 08/773,475 filed December 23, 1996.

15

Background of the Invention

Chemical reactions in biological systems are almost always facilitated by the action of one or more catalysts. Enzymes, which are proteins that catalyze biological reactions, are known for their catalytic efficiency and specificity. Enzymes typically accelerate reactions by factors of 1 million or more. Many reactions in biological systems do not occur at perceptible rates in the absence of enzymes.

Enzymes are highly specific in the type of reaction catalyzed as well as in the particular substrates which are acted upon. One broad category of enzymes includes the proteolytic enzymes which catalyze the hydrolysis of peptide bonds. Proteolytic enzymes, also known as proteases, vary significantly in their degree of specificity. For example, subtilisin, which comes from certain bacteria, will cleave peptide bonds regardless of the nature of the side chains adjacent to the bond. Trypsin is quite specific in that it splits peptide bonds on the carboxyl side of lysine and arginine residues only. Thrombin, an enzyme participating in blood clotting, is even more specific than trypsin. Thrombin only cleaves between arginine and glycine residues. These are only a very few examples of proteases; many other proteases are known. There are several general categories of proteases. These categories include serine, cysteine, aspartic, and metalloproteases. This

5 classification is based on the most prominent functional group at the active site of the proteases. The serine proteases are of particular interest relative to the current invention.

Much information now exists on the molecular structure and function of many serine proteases from diverse species. The majority of these enzymes consist of a single polypeptide chain of molecular weight 25,000-30,000. Chymotrypsin and subtilisin are both  
10 members of the serine protease family. Like other proteases, serine proteases cleave peptide bonds within a polypeptide to produce two smaller peptides. The cleavage reaction will typically proceed through an intermediate transition state which is facilitated by the presence of the protease. For serine proteases, the formation of an acyl-enzyme intermediate involving a reactive serine residue is the first step in the hydrolysis reaction. Deacylation of the acyl-  
15 enzyme intermediate is the second step in the hydrolysis. Like other proteases, serine proteases achieve their catalytic activity by lowering the activation energy for a specific hydrolysis reaction.

Proteases can be obtained from a wide variety of sources including fungi, bacteria, and eukaryotic cells. Although proteases have been obtained from many bacteria, relatively  
20 few proteases have been identified from bacteria which are known to live in extremely hot environments. Bacteria capable of growing at or above 80°C-100°C are generally known as extreme thermophiles or hyperthermophiles. Such highly thermophilic microorganisms have been the object of considerable scrutiny by researchers attempting to gain insight into the biochemical mechanism which enables these microbes to survive under such extreme  
25 conditions.

A number of microorganisms have been isolated from extremely hot environments. These microorganisms have been studied and certain useful compounds have been identified. For example, thermostable DNA polymerases have been obtained from *Thermus aquaticus*. Proteases have been isolated from thermophiles including *T. aquaticus*, *Desulfurococcus*  
30 species, *Pyrococcus furiosus*, *Sulfolobus acidocaldarius*, *Thermococcus stetteri*, and *Pyrobaculum aerophilum*. However, difficulties in culturing extremophiles have limited the number of these microbes which have been characterized as well as the number of useful

5 compounds isolated therefrom (Brennan, *Chemical and Engineering News*, October 14, 1996).

Stetter, *et al.* identified microorganisms from the hot springs of Vulcano Island, Italy, that flourish at temperatures exceeding 100°C (Stetter, K.O. "Microbial Life in Hyperthermal Environments," *ASM News* 61:285-290, 1995; Stetter, K.O., Fiala, G., Huber, R. And  
10 Segerer, A. "Hyperthermophilic Microorganisms," *FEMS Microbiol. Rev.* 75:117-124, 1990). While thermophilic organisms that grow optimally at 60°C have been known for many years, the hyperthermophilic (or extremely thermophilic) microorganisms belong to a new evolutionary class called Archaea (Woese, C.R., Kandler, O. and Wheelis, M.L. "Towards a Natural System of Organisms: Proposal for the Domains Archaea, Bacteria, and  
15 Eucarya," *Proc. Natl. Acad. Sci. USA* 87:4576-4579, 1990). The Archaea are believed to have originated over a billion years ago during the epoch when the Earth was cooling. Consequently their evolutionary development was set in motion within the environment of hot springs and deep sea hydrothermal vents. One member of this new group is *Pyrococcus furiosus* which grows optimally at 100°C-110°C (Fiala, G. and Stetter, K.O. "*Pyrococcus furiosus* s. Nev. Represents a Novel Genus of Marine Heterotrophic Archaeobacteria Growing  
20 Optimally at 100°C," *Arch. Microbiol.* 145:56-61, 1986). *Pyrococcus furiosus* is an obligate heterotroph that can be grown on polymeric substrates including protein and starch at temperatures of up to about 103°C. Preparations containing proteolytic enzymes prepared from *Pyrococcus furiosus* have been previously described in U.S. Patent Nos. 5,242,817 and  
25 5,391,489. These patents do not describe the enzymes identified by the current applicant. Other publications describing proteases from *P. furiosus* also do not describe the current enzymes. See, for example, Blumentals, Ilse I., Robinson, Anne S., and Kelly, Robert M., "Characterization of Sodium Dodecyl Sulfate-Resistant Proteolytic Activity in the Hyperthermophilic Archaeobacterium *Pyrococcus furiosus*." *Applied and Environmental  
30 Microbiology*, 56,7:1992-1998, (1990); Eggen, Rik, Geerling, Ans, Watts, Jennifer and de Vos, Willem M., "Characterization of pyrolysin, a hyperthermoactive serine protease from the archaeobacterium *Pyrococcus furiosus*." *FEMS Microbiology Letters*, 71:17-20 (1990);

- 5 Voorhorst, Wilfried G.B., Eggen, Rik I.L., Geerling, Ans C.M., Platteeuw, Christ, Siezen, Roland J., de Vos, Willem M., "Isolation and Characterization of the Hyperthermostable Serine Protease, Pyrolysin, and Its Gene from the Hyperthermophilic Archaeon *Pyrococcus furiosus*." *Journal of Biological Chemistry*, 271,34: 20426-20431 (1996).

The use of proteolytic enzymes for selective peptide bond synthesis has been  
10 previously investigated. The majority of studies so far on protease-mediated peptide synthesis have utilized what has been called "semi-synthesis". In these reactions, the acyl donor is a substrate for the enzyme (amide or ester). The substrate is utilized to acylate the enzyme (e.g., a serine or thiol protease) followed by deacylation by C-terminally blocked amino acid or peptide. (See Nakatsuka, T., Sasaki, T., and Kaiser E.T. "Peptide Segment  
15 Coupling Catalyzed by the Semisynthetic Enzyme Thiolsubtilisin." *J. Am. Chem Soc.* 109:3808-3810, 1987; Abrahmsen, L., Tom, J., Burnier, J., Butsher, K.A., Kossiakoff, A., and Wells, J.A., "Engineering Subtilisin and its Substrates for Efficient Ligation of Peptide Bonds in Aqueous Solution." *Biochemistry* 30:4151-4159, 1991; Christenen, U., Drohse, H.B., and Molgaard, L., "Mechanism of Carboxypeptidase-Y-catalyzed Peptide  
20 Semisynthesis" *Eur J. Biochem.*, 210:467-473, 1992.

The ability to synthesize peptides and ligate polypeptides in aqueous solution under controlled conditions would be highly advantageous. Current protein synthesis methodologies result in much reactant and solvent toxic waste, which must be disposed of.

## 25 Brief Summary of the Invention

In one embodiment, the subject invention provides new proteases useful in the efficient hydrolysis of peptide bonds. Advantageously, these proteases have been found to be active both as endo- and exopeptidases. Therefore, these enzymes can be used in a wide  
30 variety of applications where it is needed to remove amino acids from the end of a polypeptide, or cleave the polypeptide at an internal site.

5           In a preferred embodiment, the proteases of the subject invention have a molecular weight of about 81 kD and are serine proteases which retain enzymatic activity at about 100°C. In a specific embodiment a protease of the subject invention can be obtained from the extreme thermophile *Pyrococcus furiosus*.

          A further embodiment of the subject invention concerns nucleotide sequences which  
10    encode the proteases of the subject invention. These sequences, which can be obtained from, for example, *P. furiosus*, can be used to express the enzymes of the subject invention. These sequences, and portions thereof, are also useful as nucleotide probes to identify and characterize other related sequences. The nucleotide sequences of the subject invention can also be used as primers in PCR procedures used to obtain or characterize additional  
15    nucleotide sequences of the subject invention.

          A further aspect of the subject invention concerns antibodies to the proteases described herein. These antibodies can be used to identify and/or characterize the proteases of the subject invention.

          A further aspect of the subject invention pertains to the use of the proteases described  
20    herein in polypeptide synthesis procedures. These enzymes can be used to facilitate highly specific and efficient peptide synthesis. The enzymes of the subject invention can be used to ligate two or more peptides (reversal of endopeptidase activity), or successively add single amino acids to a peptide chain (reversal of carboxypeptidase activity). The enzymes of the subject invention can be used to synthesize peptide bonds at high temperatures with high  
25    yields. The synthesis of peptide bonds occurs, according to the subject invention, at equilibrium. The enzyme catalyzed peptide syntheses according to the subject invention are stereospecific, require little if any side chain protection and are devoid of racemization problems. Also, the ability to carry out these reactions in an aqueous solution is advantageous compared to current peptide synthesis procedures which result in the  
30    production of substantial quantities of solvent toxic waste.

          A further aspect of the subject invention concerns methods for identifying thermostable proteases. These methods involve the identification of the formation of protein

5 or peptide synthesis products produced by the ligation of substrates when a composition containing these known substrates is heated. The formation of polypeptides from the known substrates is indicative of the thermostable proteases present in the mixture.

### Detailed Disclosure of the Invention

10 In one embodiment, the subject invention pertains to novel serine proteases which can be obtained from extremely thermophilic microorganisms. The enzymes of the subject invention are catalytically active at temperatures above 60°C and, therefore, are useful in a variety of industrial processes.

Specifically exemplified herein is a novel serine protease which can be obtained from  
15 the extreme thermophile *Pyrococcus furiosus*. This enzyme has an apparent molecular weight of about 81 kDa as determined by SDS gel electrophoresis. Those skilled in the art will recognize that the apparent molecular weight of a protein as determined by gel electrophoresis will sometimes differ from the true molecular weight. Therefore, reference herein to the 81 kDa enzyme of the subject invention is understood to refer to proteins which  
20 migrate on a gel, as described herein, in a manner which is consistent with a protein of approximately that size, even if the true molecular weight is somewhat different.

The serine protease specifically exemplified herein is a carboxypeptidase enzyme. Thus, it belongs to the class of enzymes known as serine carboxypeptidases. The exemplified enzyme can act as an amidase, anilidase, and esterase. The enzyme recognizes  
25 both arginine and aromatic residues such as phenylalanine in the P1 position (nomenclature of Schechter and Berger)(Schechter, I., and Berger, A. "On the Size of the Active Site in Proteases. I. Papain." *Biochem. Biophys. Res. Commun.* 27:157-162, 1967). The enzyme is also an endopeptidase since it yields prophe + argpNA from PPANA (D-pro-phe-arg-pNA).

Certain of the properties of the serine protease specifically exemplified herein are  
30 very unique: 1) the enzyme is both an endopeptidase as well as a carboxypeptidase, 2) the enzyme displays intense product inhibition toward several synthetic peptide substrates, and 3) it is able to catalyze high-yield peptide synthesis.

5           The broad proteolytic activity of the enzymes of the subject invention as well as their thermal stability make these enzymes useful in a variety of protease applications. The high temperature proteolysis carried out using the enzymes of the subject invention is useful for many industrial applications including the food processing industry and waste removal.

          The enzymes of the subject invention can also be used in peptide and protein  
10   synthesis. For this use, peptides (or polypeptides) can be efficiently joined in the presence of the enzymes of the subject invention by increasing the temperature of the reaction mixture until the thermodynamics favor the formation of peptide bonds and, thus, the synthesis of a longer polypeptide from peptide fragments. This use of the enzymes of the subject invention is made possible by the enzymes' retention of enzymatic activity at elevated temperatures.

15           Thus, in addition to their utility as proteases, the enzymes of the subject invention are capable of synthesizing peptide bonds with high yields. The utilization of these enzymes in protein synthesis has many advantages over current protein synthesis methods, which are based on semi-synthesis. One of the major practical problems associated with "semi-synthesis" is that it must be kinetically monitored, or controlled. That is, the synthetic  
20   reaction must be terminated at or near the time when synthetic yield is at a maximum. Otherwise, proteolysis of the synthetic product will supervene and it will be driven thermodynamically to essentially complete hydrolysis. Equilibrium peptide synthesis according to the subject invention does not suffer this disadvantage. Also, use of these enzymes in protein synthesis is particularly advantageous because stereospecificity is  
25   preserved. Furthermore, group protection and toxic solvents are unnecessary when polypeptide synthesis is carried out according to the subject invention. Unlike previously known procedures, the peptide synthesis carried out according to the subject invention can be done without the use of harmful organic solvents.

          The subject invention further provides methods for identifying thermostable enzymes.  
30   In one embodiment crude cellular preparations (or other compositions which may contain a thermostable enzyme) can be assayed for the presence of thermostable enzymes. In this embodiment, peptide and/or polypeptide substrates can be added to the crude preparation.

5 The composition can then be heated and analyzed for the presence of ligated peptides or polypeptides. In this embodiment, thermostable enzymes will catalyze the synthesis of polypeptides from the peptide or polypeptide substrates. Thus, the presence of thermostable enzymes can be identified by the formation of ligated polypeptides after heat treatment. The enzyme(s) responsible for the activity can then be identified through sequential isolation  
10 steps which remove inactive compounds and result in the isolation of the thermostable enzymes. The enzymes can then be purified and characterized according to standard procedures. The subject invention includes the enzymes obtained according to this assay procedure.

The new proteins provided here are defined according to several parameters. One  
15 critical characteristic of the proteins described herein is thermostable enzymatic activity. In a specific embodiment, these proteins are serine proteases. The enzymes and genes of the subject invention can be further defined by their amino acid and nucleotide sequences. The sequences of the molecules can be defined in terms of homology to certain exemplified sequences as well as in terms of the ability to hybridize with certain exemplified sequences.  
20 The enzymes provided herein can also be identified based on their immunoreactivity with certain antibodies.

The polynucleotide sequences and enzymes useful according to the subject invention include not only the full length sequences disclosed herein but also fragments of these sequences, as well as variants, mutants, and fusion proteins which retain the characteristic  
25 enzymatic activity of the proteins specifically exemplified herein. As used herein, the terms "variants" or "variations" of genes refer to nucleotide sequences which encode the same enzyme or which encode equivalent enzymes having proteolytic activity. As used herein, the term "equivalent enzymes" refers to enzymes having the same or essentially the same biological activity as the exemplified enzymes, albeit with different specificity.

30 It would be apparent to a person skilled in this art that genes encoding active enzymes can be identified and obtained through several means. The gene encoding the specific enzyme exemplified herein may be obtained from the specific isolate described herein. This



5 gene, or portions or variants thereof, may also be constructed synthetically, for example, by use of a gene synthesizer. Variations of genes may be readily constructed using standard techniques for making point mutations. Also, fragments of these genes can be made using commercially available exonucleases or endonucleases according to standard procedures. For example, enzymes such as *Bal31* or site-directed mutagenesis can be used to  
10 systematically cut off nucleotides from the ends of these genes. Also, genes which encode active fragments may be obtained using a variety of restriction enzymes. Proteases may be used to directly obtain active fragments of these enzymes.

Equivalent enzymes and/or genes encoding these equivalent enzymes can be derived from extreme thermophile isolates and/or DNA libraries using the teachings provided herein.  
15 There are a number of methods for obtaining the enzymes of the instant invention. For example, antibodies to the specific enzyme disclosed and claimed herein can be used to identify and isolate other such enzymes from a mixture of proteins. Specifically, antibodies may be raised to the portions of the enzyme which are most distinct from other enzymes. These antibodies can then be used to specifically identify equivalent enzymes with the  
20 characteristic activity by immunoprecipitation, enzyme linked immunosorbent assay (ELISA), or western blotting. Antibodies to the enzyme disclosed herein, or to equivalent enzymes, or fragments of these enzymes, can readily be prepared using standard procedures in this art. The genes which encode these enzymes can then be obtained from the host cell.

The subject invention concerns not only the polynucleotide sequences which encode  
25 these enzymes but also the use of these polynucleotide sequences to produce recombinant hosts which express the enzymes. The enzyme-encoding genes of the subject invention can be introduced into a wide variety of microbial or plant hosts. Expression of the gene results, directly or indirectly, in the intracellular production and maintenance of the enzyme.

Fragments and equivalents which retain the enzymatic activity of the exemplified  
30 proteins would be within the scope of the subject invention. Also, because of the redundancy of the genetic code, a variety of different DNA sequences can encode the amino acid sequences disclosed herein. It is well within the skill of a person trained in the art to create

5 these alternative DNA sequences encoding the same, or essentially the same, proteins. These variant DNA sequences are within the scope of the subject invention. As used herein, reference to "essentially the same" sequence refers to sequences which have amino acid substitutions, deletions, additions, or insertions which do not materially affect enzymatic activity. Fragments retaining enzymatic activity are also included in this definition.

10 A further method for identifying the proteins and genes of the subject invention is through the use of oligonucleotide probes. These probes are detectable nucleotide sequences. These sequences may be detectable by virtue of an appropriate label or may be made inherently fluorescent as described in International Application No. WO93/16094. As is well known in the art, if the probe molecule and nucleic acid sample hybridize by forming a  
15 strong bond between the two molecules, it can be reasonably assumed that the probe and sample have substantial homology. Preferably, hybridization is conducted under stringent conditions by techniques well-known in the art, as described, for example, in Keller, G.H., M.M. Manak (1987) *DNA Probes*, Stockton Press, New York, NY., pp. 169-170.

As used herein "stringent" conditions for hybridization refers to conditions which are  
20 able to distinguish genes encoding heat stable serine proteases from unrelated genes. Specifically, hybridization of immobilized DNA on Southern blots with <sup>32</sup>P-labeled gene-specific probes can be performed by standard methods (Maniatis *et al.*). For double-stranded DNA gene probes, hybridization can be carried out overnight at 20-25° C below the melting temperature (T<sub>m</sub>) of the DNA hybrid in 6X SSPE, 5X Denhardt's solution, 0.1% SDS, 0.1  
25 mg/ml denatured DNA. The melting temperature is described by the following formula (Beltz, G.A., K.A. Jacobs, T.H. Eickbush, P.T. Cherbas, and F.C. Kafatos [1983] *Methods of Enzymology*, R. Wu, L. Grossman and K. Moldave [eds.] Academic Press, New York 100:266-285).

$$T_m = 81.5^{\circ} \text{C} + 16.6 \log[\text{Na}^+] + 0.41(\% \text{G} + \text{C}) - 0.61(\% \text{formamide}) - 600/\text{length of duplex}$$
  
30 in base pairs.

Washes are typically carried out as follows:

- 5           (1)     Twice at room temperature for 15 minutes in 1X SSPE, 0.1% SDS (low stringency wash).
- (2)     Once at  $T_m - 20^\circ\text{C}$  for 15 minutes in 0.2X SSPE, 0.1% SDS (moderate stringency wash).

For oligonucleotide probes, hybridization can be carried out overnight at  $10-20^\circ\text{C}$   
10 below the melting temperature ( $T_m$ ) of the hybrid in 6X SSPE, 5X Denhardt's solution, 0.1% SDS, 0.1 mg/ml denatured DNA.  $T_m$  for oligonucleotide probes can be determined by the following formula:

$$T_m (^{\circ}\text{C}) = 2(\text{number T/A base pairs}) + 4(\text{number G/C base pairs})$$

(Suggs, S.V., T. Miyake, E.H. Kawashime, M.J. Johnson, K. Itakura, and R.B. Wallace  
15 [1981] *ICN-UCLA Symp. Dev. Biol. Using Purified Genes*, D.D. Brown [ed.], Academic Press, New York, 23:683-693).

Washes can be typically carried out as follows:

- (1)     Twice at room temperature for 15 minutes 1X SSPE, 0.1% SDS (low stringency wash).
- 20           (2)     Once at the hybridization temperature for 15 minutes in 1X SSPE, 0.1% SDS (moderate stringency wash).

With the teachings provided herein, one skilled in the art could readily produce and use the various enzymes and polynucleotide sequences of the novel enzymes described herein.

25           Detection of the probe provides a means for determining in a known manner whether hybridization has occurred. Such a probe analysis provides a rapid method for identifying enzyme-encoding genes of the subject invention. The nucleotide segments which are used as probes according to the invention can be synthesized using a DNA synthesizer and standard procedures. These nucleotide sequences can also be used as PCR primers to amplify genes  
30 of the subject invention.

Certain enzymes of the subject invention have been specifically exemplified herein. Since these enzymes are merely exemplary of the enzymes of the subject invention, it should

5 be readily apparent that the subject invention comprises variant or equivalent enzymes (and  
nucleotide sequences coding for equivalent enzymes) having the same or similar enzymatic  
activity of the exemplified serine protease. Equivalent enzymes will have amino acid  
homology with the exemplified enzyme. This amino acid homology will typically be greater  
than 60%, preferably be greater than 75%, more preferably greater than 80%, more preferably  
10 greater than 90%, and can be greater than 95%. The amino acid homology will be highest in  
critical regions of the enzyme which account for biological activity or are involved in the  
determination of three-dimensional configuration which ultimately is responsible for the  
biological activity. In this regard, certain amino acid substitutions are acceptable and can be  
expected if these substitutions are in regions which are not critical to activity or are  
15 conservative amino acid substitutions which do not affect the three-dimensional  
configuration of the molecule. For example, amino acids may be placed in the following  
classes: non-polar, uncharged polar, basic, and acidic. Conservative substitutions whereby  
an amino acid of one class is replaced with another amino acid of the same type fall within  
the scope of the subject invention so long as the substitution does not materially alter the  
20 biological activity of the compound. Table 1 provides a listing of examples of amino acids  
belonging to each class.

---

**Table 1.**

---

Class of Amino Acid	Examples of Amino Acids
Nonpolar	Ala, Val, Leu, Ile, Pro, Met, Phe, Trp
Uncharged Polar	Gly, Ser, Thr, Cys, Tyr, Asn, Gln
Acidic	Asp, Glu
Basic	Lys, Arg, His

---

5           In some instances, non-conservative substitutions can also be made. The critical factor is that these substitutions must not significantly detract from the biological activity of the enzyme.

          Following are examples which illustrate procedures for practicing the invention.  
10   These examples should not be construed as limiting. All percentages are by weight and all solvent mixture proportions are by volume unless otherwise noted.

Example 1 — Isolation of a Serine Protease of the Subject Invention

*Pyrococcus furiosus* cells were obtained from Professor C.B. Anfinsen's laboratory at  
15   Johns Hopkins. These cells can also be obtained from Dr. Michael W.W. Adams at the University of Georgia. Natural Sources of these cells have been described in the literature. See, for example, references cited in the Background of the invention. Cells are centrifuged and 100 g (wet weight) is suspended in water and sonicated. Cell debris is removed by centrifugation and the resulting supernatant solution is dialyzed against 0.05 M tris-HCl, pH  
20   7.5.

          Four chromatographic steps are utilized to yield an electrophoretically and ultracentrifugally homogeneous macromolecule. The entire purification scheme utilizes FPLC (Pharmacia) and all chromatographic fractions are screened for enzyme activity (at 25°C) using N-benzoyl-L-arginine-p-nitroanilide (BAPNA) as substrate.

25           The chromatographic steps are as follows:

- a.       The crude sonicate is applied to Mono Q 10/10 equilibrated with 0.05 M tris-HCl, pH 7.5 linear 184 ml gradient to 0.4 M KCl, 3 ml/min. Active fractions are pooled and dialyzed vs. 0.05 M sodium acetate, pH 4.5.
- b.       Mono S 10/10, 0.05 M sodium acetate, pH 4.5, linear 368 ml gradient to 1.0  
30       M NaCl at 3 ml/min.
- c.       Mono Q 5/5 0.05 M histidine, pH 5.5, linear gradient (23 ml) to 0.4 M KCl at a flow rate of 1 ml/min.

- 5           d.       The pooled fractions from (c) are concentrated by membrane filtration and applied to a 2.6 x 60 cm column of Superdex-200 equilibrated with 0.1 M tris-HCl, 0.1 NaCl, pH 7.5.

Throughout the above ion-exchange steps (a-c), the chromatographic profiles reveal 2 prominent protein peaks that display BAPNA anilidase activity, together with other BAPNA-positive peaks in much lower amounts that are successively eliminated with each column step. The last step (gel filtration) yields 2 well-separated protein fractions that represent approximately 80% and 15% of the anilidase activity present in the original cell sonicate. The most abundant of these 2 proteins is the one used for all of the studies described below. It emerges from the Superdex-200 column with an apparent molecular weight of about 110,000 as judged from its partition coefficient determined with standard gel filtration molecular weight marker proteins. Polyacrylamide gel electrophoresis (SDS-PAGE) yields a single sharp band under reducing conditions. The estimated molecular weight of this band is approximately 81,000. The yield is approximately 1 mg of pure protein from 100 g wet cells.

20

Example 2 – Extinction Coefficient, Apparent Partial Specific Volume, and Molecular Weight of Serine Protease of the Subject Invention

The protein was hydrolyzed (constant boiling HCl) for 18, 22, 24 and 26 hours. From absorbance measurements (280 nm) and the methods of Edelhoch (Edelhoch, H. “Spectroscopic Determination of Tryptophan and Tyrosine in Proteins.” *Biochemistry* 6:1948-1954, 1967), the extinction coefficient was calculated to be  $1.31 \text{ ml mg}^{-1} \text{ cm}^{-1}$ .

Sedimentation equilibrium measurements utilized a Beckman Model E ultracentrifuge equipped with a split-beam scanner and multiplexer for visualization of two centrifuge cells during the same run. The high speed method of Yphantis (Yphantis, D.A. “equilibrium Ultracentrifugation of Dilute Solutions,” *Biochemistry* 3:294-303, 1964) was employed together with the methods of Edelstein and Schachman for simultaneous measurement of the partial specific volume (Edelstein, S.J. and Schachman, H.K. “The

5 Simultaneous Determination of Partial Specific Volumes and Molecular Weights with  
Microgram Quantities.” *J. Biol. Chem.* 242:306-311, 1967). One cell contained protein  
dialyzed thoroughly against 0.1 M tris-HCl, pH 7.5, in H<sub>2</sub>O and the second cell contained  
the enzyme in the same buffer with 99% D<sub>2</sub>O as solvent (densities of the buffer solutions  
were measured pycnometrically). Centrifugation (20,000 RPM, 23.5EC) yielded a molecular  
10 weight of approximately 81,500 and an apparent partial specific volume (Casassa, E.F. and  
Eisenberg, H. “Thermodynamic Analysis of Multicomponent Solutions.” *Adv. Prot. Chem.*  
19:287-393, 1964) of 0.789 ml/g. This is a surprisingly high value for the specific volume of  
a protein and it implies a larger than expected Stokes radius, which may explain why the  
protein emerges earlier upon gel filtration than would be anticipated for a protein of  
15 molecular weight of 81,500. Plots of  $\ln c$  vs  $r^2$  were strictly linear—a feature that indicates  
size homogeneity. The close similarity of the molecular weight to that obtained by SDS-  
PAGE indicates that the protein has a single polypeptide chain structure.

### Example 3 – Stability of Enzymic Activity at High Temperature

20 For all kinetic experiments at high temperatures, sodium phosphate (0.025 M) was  
used as a buffer. The temperature coefficient of this buffer is so small that slight changes in  
pH with temperature do not significantly affect the kinetic data.

To assess stability of the enzyme at high temperature, a solution of the protein in the  
above buffer, pH 7.0, was incubated at 82.0 +/- .05°C. Aliquots were removed at hourly  
25 intervals up to 8 hr, and initial velocities were measured (BAPNA as substrate, Varian 2290  
recording spectrophotometer) at 25.0° +/- .05°C (Erlanger, B.F., Kokowski, N. and Cohen,  
W. “The Preparation and Properties of Two New Chromogenic Substrates of Trypsin.” *Arch.*  
*Biochem. Biophys.* 95:271-278, 1961). No decrease in enzyme activity was observed over  
this time period.

30

#### 5 Example 4 – Activation Enthalpy and Entropy as a Function of Temperature

To explore the basis for the extremely large increases in enzyme activity with temperature, initial velocity ( $v_o$ ) measurements were made at 6 different substrate concentrations in order to determine  $k_{cat}$  and  $K_m$  at a series of temperatures in the range of 25.0 to 83.5°C. Temperature was controlled to within  $\pm 0.5^\circ\text{C}$  with a large water bath and  
 10 jacketed cuvettes. Two substrates were used for these experiments - BAPNA and D-pro-phe-arg-pNA(PPANA). From these data, Arrhenius plots were constructed ( $\ln v_o$  vs  $1/T^\circ\text{K}$ ).

Within the temperature range accessible to mesophilic enzymes, such plots are linear. The plot with BAPNA as substrate exhibits considerable continuous curvature. The data was fitted to a second degree polynomial and the slopes determined as a function of  
 15 temperature to compute Arrhenius activation energies. These values were used to compute activation enthalpy ( $\Delta H^*$ ) and entropy ( $\Delta S^*$ ) as a function of temperature (Eyring, H. (1963) *Modern Chemical Kinetics*, Rheinhold, New York). Table 2 summarizes these values for BAPNA and D-pro-phe-arg-pNA at 25.0° and 80.0°C.

**Table 2.**

$\Delta H^*$  and  $\Delta S^*$  as a Function of Temperature

Substrate	T°C	$\Delta H^*$ (kcal/mole)	$\Delta S^*$ (e.u)
BAPNA	25.0	17.2	2.2
BAPNA	80.0	3.1	-41.7
PPANA	25.0	26.1	23.8
PPANA	80.0	15.4	-9.0

20



- 5           The data and Table 2 show that the activation energy barrier falls markedly, in a continuous fashion, as temperature increases. Furthermore, the fact that the Arrhenius plots are smoothly monotonic indicates that no sharp, discontinuous structural transition (*e.g.*, from less active to more active enzyme) occurs. The process appears to be continuous from 25-80°C. The entropies become negative at higher temperatures.

10

#### Example 5 – Active Site Inhibitor Studies

To assign the class of protease to which the *P. furiosus* protease belongs, several reagents were used to identify catalytically important residues. BAPNA was the substrate for all of the following results:

- 15           a.     Incubation of the enzyme with  $10^{-3}$  M EDTA for 24 hr had no effect upon enzyme activity.
- b.     One of the active-site inhibitors of mesophilic serine proteases is phe-pro-argCH<sub>2</sub>Cl. This reagent reacts rapidly and specifically with the active site histidine residue of the active site triad: ser-his-asp (Shaw, E. "Site-Specific
- 20           Reagents for Chymotrypsin and Trypsin." *Math. Ensemble*. 11:677-686, 1967). Treatment of the *P. furiosus* enzyme with this inhibitor ( $1.98 \times 10^{-6}$  M) together with kinetic measurements of BAPNA enzymic activity (BAPNA =  $3.8 \times 10^{-4}$  M) vs time at 25.0°C yielded a pseudo first order inactivation rate constant =  $0.022 \text{ sec}^{-1}$ .
- 25           c.     Treatment of the enzyme with PMSF (phenyl-methyl-sulfonyl fluoride)  $8 \times 10^3$  molar excess over enzyme, resulted in complete inhibition of enzyme activity with BAPNA as the assay substrate at 25.0°C.
- d.     The following reagents had no affect whatsoever upon the activity of this enzyme at 25.0°C: i) 0.025 M sodium phosphate, pH 7.0; ii) iodoacetamide:
- 30           100-fold molar excess: enzyme; iii) bisdithiodintrobenzene (Ellman's reagent).

- 5           Taken together, the above results confirm that the protease active site does indeed contain histidine and serine, and that it is a serine protease, not a thiol enzyme.

#### Example 6 – Kinetic Studies and Substrate Specificity

10           The hydrolysis kinetics of various synthetic peptide and ester protease substrates were studied. Two different methods have been used to analyze the kinetic properties of the enzyme. The first utilized initial velocity data at several substrate concentrations with analysis by plotting  $s/v$  vs  $s$  (Hanes plot). This method was always used when it was possible that the protease could cleave at more than one site (*e.g.*, PPANA). The second method utilizes all data of the progress curve of the reaction. Progress curves were analyzed in two  
15   ways:

a.        $v$  vs  $s$ .

          Instantaneous point by point derivatives were calculated from the absorbance vs time data by fitting 10 points spaced 1 second apart by least squares to a second degree polynomial. Derivatives were computed analytically and were used to construct plots of  $v$  vs  
20    $s$ , which were then analyzed by non-linear regression to obtain  $K_m$ ,  $k_{cat}$  and any product inhibition constant ( $K_i$ ) if it exists. The advantages of this method are that it does not require a precise value of the initial substrate concentration, (Koerber, S.C. and Fink, A.L. "The Analysis of Enzyme Progress Curves by Numerical Differentiation, Including Competitive Product Inhibition and Enzyme Reactivation." *Anal. Biochem.* 165:75-87, 1987), and it can  
25   readily be used to correct for non-enzymic substrate hydrolysis at high temperatures. For all measurements at elevated temperature, the first-order rate constants for non-enzymic hydrolysis were determined for all substrates. The analytic method described here was used both for reactions at 25.0°C and 80.0°C.

b.       Integrated rate equation analyses:  $t$  vs Absorbance.

30           For these analyses, the integrated Michaelis-Menten equation was used, but not in linearized form, for reasons relating to product inhibition. Plots of time (now the dependent variable) vs absorbance were constructed from the raw data and analyzed by non-linear

5 regression. In these analyses, the initial substrate concentration is taken as a parameter to be fit in the non-linear regression—a procedure that has been shown to provide more accurate values of the kinetic parameters (Newman, P.F.J., Atkins, G.L. and Nimmo, I.A., “The Effect of Systematic Error on the Accuracy of Michaelis Constants and Maximum Velocities Estimated by Using the Integrated Michaelis-Menten Equation.” *Biochem. J.* 143:779-781, 10 1974). This method was used for analyses at 25.0°C only.

c. Computer methods.

Subroutines for the above analyses were written to analyze the kinetic data for: product competitive inhibition, non-competitive inhibition, 2-product inhibition, and simple Michaelis-Menten kinetics. An HP 9000 series work station was used for this purpose, using 15 HP BASIC.

The substrates studied by the above methods are listed here and all data are summarized in Table 3. All kinetic studies utilized a Cary 3E double beam spectrophotometer and absorbance vs time data were collected by computer at 1 second intervals. In cases where a peptide substrate could potentially be cleaved at more than one 20 site, thin-layer chromatography was used to establish the products formed (silica gel; butanol: acetone: NH<sub>4</sub>OH:water - 37:37:19:7).

1. BAPNA (Benzoyl-arginine-p-nitroanilide)

The slope of the Hanes plot for BAPNA at 25.0° is negative. This can only mean that the reaction is product inhibited and that  $K_i < K_m$  for this substrate.

25 Table 3 shows that  $K_i$  is 10 times less than  $K_m$ . This result reflects very tight binding of one of the two products of BAPNA hydrolysis to the protein. To determine the inhibition constants of the products, initial rate measurements of BAPNA were conducted with varying concentrations of the two products - benzoyl-arg and p-nitroaniline. p-nitroaniline displayed no detectable inhibition at concentrations as high as  $5 \times 10^{-4}$  M, whereas benzoyl-L-arginine 30 strongly inhibited with  $K_i = 6.50 \times 10^{-5}$  M. This number is close to that ( $7.08 \times 10^{-5}$  M) obtained from total progress curve analyses. This close agreement of the  $K_i$  values determined by totally different analytic methods also lends credence to results from the total

5 progress curve data that have been analyzed by non-linear regression methods. Such strong product inhibition is exceptional.

A second unusual feature concerns the  $k_{\text{cat}}$  value for BAPNA at 25.0°C (see Table 3). The  $k_{\text{cat}}$  value for BAPNA as substrate is greater than that for trypsin under the same conditions ( $k_{\text{cat}} = 2.7 \text{ sec}^{-1}$ ). The value of  $k_{\text{cat}}$  is 65 times greater at 80°C than at 25°C.

10 2. PPANA (D-pro-phe-arg-pNA)

Continuous progress curve analyses of this substrate indicated that it can be cleaved at more than one bond. The substrate was completely depleted before the expected amount of pNA had been liberated. Therefore, initial rates were used to determine the values given in Table 3. It will be seen that  $k_{\text{cat}}$  is 120 times greater at 80°C than at 25°C. TLC  
15 demonstrated that the protease:

- a. produces pro-phe-arg + pNA
- b. cleaves pro-phe-arg 6 pro-phe + arg
- c. cleaves substrate 6 pro-phe + arg-pNA
- d. does not hydrolyze arg-pNA at all, even at 80°C

20 The above results indicate that this protease has endopeptidase as well as carboxypeptidase activity.

3. BTEE (benzoyl-tyrosine ethyl ester).

The protease cleaves this substrate with a large  $k_{\text{cat}}$  at 25°C. Thus, the protease is also an esterase.

25 4. FAGLA (furyl-acryloyl-glycyl-leucine amide)

This commonly used thermolysin substrate shows intense product inhibition. TLC demonstrates that the protease only deamidates the substrate. It does not split the gly-leu bond.  $K_i < K_m$ .

5. FAPP (furyl-acryloyl-phe-phe)

30 Cleavage occurs to liberate only FAP plus phe (TLC), and product inhibition is not observed. Only a simple Michaelis-Menten mechanism will fit the experimental data. This is the second example of carboxypeptidase-like activity.

5           6.     FAAPA (furyl-acryloyl-ala-phe-amide)

Like FAGLA, the enzyme deamidates this substrate with  $K_i < K_m$ . TLC reveals only FAAP as a product. No cleavage occurs at the ala-phe bond.

7.     ZPA (carbobenzoxy-phe-ala)

Ala is liberated in carboxypeptidase fashion, again with  $K_i < K_m$ .

10          8.     FAPGG (furyl-acryloyl-phe-gly-gly)

The enzyme only hydrolyzes the phe-gly bond with liberation of FAP + gly- gly. No product inhibition is observed. This is another example of the endopeptidase activity of the enzyme.

15          9.     Finally, two reagents remain unaltered by the enzyme—HA (hippuryl-arginine)  
and ZGP (carbobenzoxy-gly-phe).

5

**Table 3.**

Substrate	T°C	$k_{cat}(\text{sec}^{-1})$	$K_m(\text{M})$	$K_i(\text{M})$
BAPNA	25.0	6.27	$7.37 \times 10^{-4}$	$7.08 \times 10^{-5}$
BAPNA	80.0	406	$7.78 \times 10^{-4}$	$2.20 \times 10^{-4}$
PPANA	25.0	.045	$4.77 \times 10^{-4}$	N.D.
PPANA	80.0	5.44	$4.99 \times 10^{-4}$	N.D.
BTEE	25.0	27.4	$6.57 \times 10^{-4}$	$5.6 \times 10^{-3}$
FAGLA	25.0	0.67	$6.08 \times 10^{-6}$	$2.79 \times 10^{-6}$
FAPP	25.0	4.58	$3.75 \times 10^{-4}$	None
FAAPA	25.0	7.31	$3.86 \times 10^{-5}$	$1.57 \times 10^{-5}$
HA	25.0	---	---	---
ZPA	25.0	3.65	$5.55 \times 10^{-4}$	$1.34 \times 10^{-4}$
ZGP	25.0	---	---	---
FAPGG	25.0	10.64	$4.15 \times 10^{-4}$	---

All of the above are mean values from at least 4 separate experiments. Except for PPANA, values are means of  $v$  vs  $s$  and integrated rate equation analyses. PPANA values were obtained from initial velocity measurements, analyzed by the Hanes equation. Solvent: 0.025 M Na Phosphate, pH 7.0. 2000 - 5000 data points were used for the above analyses (except for PPANA). N.D. means not determined.

- 5 All peptide substrates listed in Table 3 yield product inhibition constants ( $K_i$ ) that are approximately 3-10 lower than  $K_m$  values.

#### Example 7 – Substrate Specificity

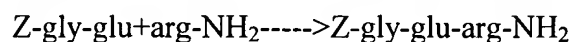
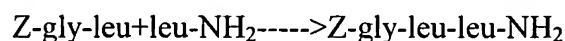
- 10 Treatment of the insulin B-chain with the serine protease of the subject invention (80°C, phosphate buffer, pH 7.0) for 1 hour yields 5 ninhydrin-positive components by TLC. No non-enzyme hydrolysis of substrate was detected after this time period. One of the new components is alanine which the carboxy terminus of the B-chain. None of the others is a free amino acid. These studies with insulin provide further evidence that the enzyme is an endopeptidase.

15

#### Example 8 – Peptide Synthesis

- Benzoyl-arginine ethyl ester was selected as an acyl donor because of the very low product inhibition constants observed for benzoyl-arginine ( $K_i = 7.08 \times 10^{-5}$  M, Table 3). Leucine-amide was chosen as nucleophile. Reaction solutions 0.025 M Na-phosphate, pH 20 7.0, at 25.0°C contained  $6.9 \times 10^{-3}$  M BAEE, 0.091 M leucine-amide and 0.015 mg/ml protease. Aliquots were removed over time and analyzed by TLC as described earlier. After 45 min at 25°C, the yield of benzoyl-arginine-leucine amide was approximately 50% from TLC and HPLC analyses. This represents considerable synthesis over a relatively short time period and it occurs at 25°C.

- 25 Two different peptides have been synthesized at pH 5.5 and 85°C. These peptides are as follows:



- Reactions were terminated after 3 hours at 85°C and reaction solutions were analyzed by 30 FPLC. The synthetic products were verified by mass spectroscopy. For both reactions, the concentration of the amine nucleophile was 0.9M. The protease concentration was 10 µg/ml.

5 Table 4 summarizes the results. Yield is defined as the ratio of product concentration to the initial concentration of the carboxyl reactant (in percent). Theoretical yield was calculated from equations 1, 2 and 3 together with  $pK_1=3.81$  (ZGL) and 4.47 (ZGE);  $pK_2=7.93$  (leu-NH<sub>2</sub>) and 7.51 (arg-NH<sub>2</sub>).

Table 4.

Reaction	[Carboxyl]	Yield	Yield(theory)	$\Delta G^\circ_{\text{net}}$	$K_{\text{net}}(\text{M}^{-1})$
ZGL+leu-NH <sub>2</sub>	0.01M	48%	55%	-0.021	1.03
ZGE+arg-NH <sub>2</sub>	0.048M	60%	68%	-0.380	1.72

10 \* $\Delta G^\circ_{\text{net}}$  (kcal/mole) and  $K_{\text{net}}$  are the experimental values.

The yields for peptide synthesis presented in Table 4 are far greater than any previously reported values under equilibrium conditions. Moreover, the above reactions were carried out in the absence of organic co-solvents and in the absence of product precipitation to drive the reactions. As shown in Table 4, the actual yields of synthesis product are remarkably close to predicted values.

#### Example 9 - Peptide Sequencing

20 The amino acid sequence of the first 20 amino-terminal residue was found to be:  
SSIEWNEKTFAKFAYLSDPR (SEQ. ID NO. 1).

No identical matches to this sequence could be found in the data banks. An analysis of the initial eight amino acids strongly favors a single peptide chain, in agreement with electrophoretic and ultracentrifugal analyses. To obtain internal amino acid residue sequences, the protein can be digested by chymotrypsin or trypsin attached to beaded agarose. After pelleting to remove the beads, the fragmented protein can be subjected to



- 5 SDS-PAGE and stained. The digest lane can be transferred to PVDF membrane. Well-separated bands are then subjected to sequencing.

Example 10 – Cloning and Sequencing of the Gene Encoding the Serine Protease from *Pyrococcus furiosus*

- 10 The gene encoding the serine protease specifically exemplified herein can be cloned and sequenced by a person skilled in the art having the benefit of the instant disclosure. One cloning method involves screening a genomic library of *P. furiosus*. A *P. furiosus* genomic library is commercially available from Sybtrell Biotechnology. The average insert size is approximately 7 kilobase pairs indicating that approximately 8000 transformants must be  
15 screened in order to have a 99% chance of full coverage of the genome (Sambrook, J., Fritsch, E.F. and Maniatis, T. *Molecular Cloning, A Laboratory Manual*, 2nd edition. Cold Spring Harbor Laboratory Press, 1989). An oligonucleotide probe based on the partial protein sequence has been designed to identify for the serine protease gene by hybridization. This probe has the following sequence:

- 20 GA(AG)TGGAA(TC)GA(AG)AA(AG)AC (SEQ. ID NO. 2).

This oligonucleotide has only eight-fold redundancy providing a high probability of specific hybridization to the target gene DNA with minimal background.

- The library can be transformed into *E. coli* strain DH5 $\alpha$  by the CaCl<sub>2</sub> method  
25 (Sambrook, J., Fritsch, E.F. and Maniatis, T. *Molecular Cloning, A Laboratory Manual*, 2nd edition. Cold Spring Harbor Laboratory Press, 1989). The Sybtrell library was established in a pUC19 vector system so transformants can be selected based on resistance to ampicillin, and spread to yield a density of approximately 500 colonies per LB-Ap plate. Colonies can be lifted onto nitrocellulose filter discs. The discs are washed with 10% SDS to lyse the cells  
30 and the DNA denatured under alkaline conditions (0.5 M NaOH, 1.5 M NaCl, 5 min). Following neutralization and a buffer wash, the filters are dried under vacuum at 80°C for 1 hour. The probe oligonucleotide can be end labeled with <sup>32</sup>P using T4 DNA kinase (New

5 England Biolabs, Beverly, MA). The filters can be prehybridized using salmon sperm DNA as a blocking agent, and then hybridization can be at 45°C overnight. The filters are then washed under progressively higher stringency by raising the wash temperature. Positive hybridization signals can be detected by autoradiography. Colonies corresponding to positive  
10 screening until pure cultures are obtained.

The size of *P. furiosus* DNA inserts in positive clones can be determined by *EcoRI* digestion and a preliminary restriction map developed. This facilitates generation of a series of nested deletions for DNA sequence analysis. DNA sequencing can be performed in an automated DNA facility using the vector specific universal primer. Alternatively as the  
15 sequence emerges, *P. furiosus* DNA specific primers can be generated to obtain sequence from areas which are too distant from the universal priming site or simply prove difficult to sequence (*i.e.*, high GC content regions).

The following criteria can be used to establish that the cloned gene codes for the *P. furiosus* serine protease. First, the full length native protein obtained as described herein has  
20 a molecular weight of approximately 81,500 daltons suggesting an open reading frame of at least approximately 2000-2200 base pairs (taking into account some potential post-translational processing). The deduced primary sequence should match the partial amino acid sequence derived from the native protein.

It should be understood that the examples and embodiments described herein are for  
25 illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and the scope of the appended claims.